The coin tossing problem (HT21)

Problem set-up: Toss a fair coin for many times,

$\{Z_1, Z_2, \ldots\}$. $Z_i \in \{H, T\}$.

Define $\tau_1 := \inf\{n : Z_{n-1} = H, Z_n = H\}$.

$\tau_2 := \inf\{n : Z_{n-1} = H, Z_n = T\}$.

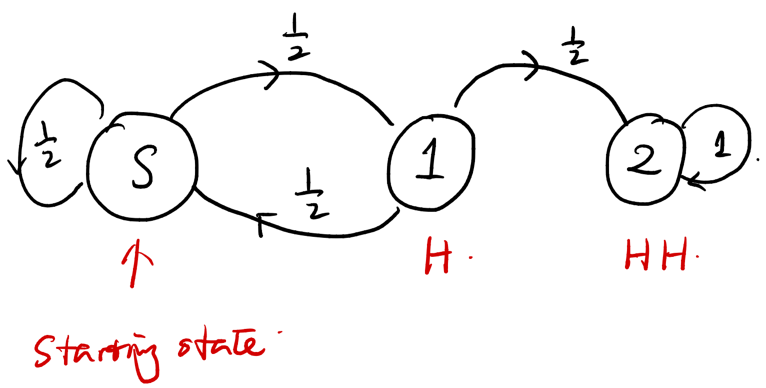
Question: Does $E[\tau_1]$ equal $E[\tau_2]$?

Intuition: $\{HT\}$ and $\{HH\}$ both happen with $\frac{1}{4}$ probability.

So $E[\tau_1] = E[\tau_2]$

Truth: $6 = E[\tau_1] \neq E[\tau_2] = 4$.

Construct a Markov chain for $\{H, H\}$.



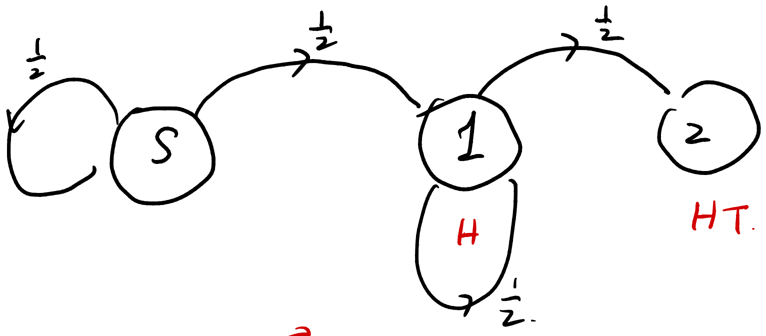
Let $s = E[\tau_{s\to 2}]$. ⟶ what we want.

$a = E[\tau_{1\to 2}]$.

"First step analysis". yields. (Expected hitting time from different states to the absorbing state).

$$\begin{cases} s = 1 + \frac{1}{2}s + \frac{1}{2}a \\ a = 1 + \frac{1}{2}s + \frac{1}{2}\times 0 \end{cases} \implies s = 6.$$

Similarly, construct a chain for $\{H, T\}$.



when fails, start from the middle of the chain.

$\implies$ shorter time of hitting.

Similarly $E[\tau_{s\to 2}] = \tilde{s}$, $E[\tau_{1\to 2}] = \tilde{a}$.

$$\begin{cases} \tilde{s} = 1 + \frac{1}{2}\tilde{s} + \frac{1}{2}\tilde{a} \\ \tilde{a} = 1 + \frac{1}{2}\tilde{a} + \frac{1}{2}\times 0 \end{cases} \implies \tilde{s} = 4$$

<u>Remark</u> Ask a question to a random person:

probability of $\{HHHH\}$ and $\{HTHT\}$ which one is bigger?

The random person probably says $\{HTHT\}$

We know that $P(\{HHHH\}) = P(\{HTHT\}) = \frac{1}{2^4}$

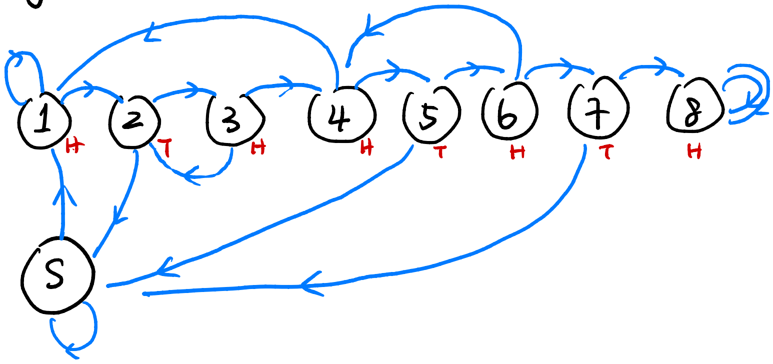
Somehow this is also correct in the sense that

it takes longer (on average) to see $\{HHHH\}$

Now another question:

What if we want to see "HTHHTHTH"?

Similarly, construct a Markov chain.



$$s = 1 + \frac{1}{2}(s+a)$$
$$a = 1 + \frac{1}{2}(a+b)$$
$$b = 1 + \frac{1}{2}(c+s)$$
$$c = 1 + \frac{1}{2}(d+b)$$
$$d = 1 + \frac{1}{2}(e+a)$$
$$e = 1 + \frac{1}{2}(f+s)$$
$$f = 1 + \frac{1}{2}(g+d)$$
$$g = 1 + \frac{1}{2}s$$

$\boxed{s = 266.}$ $\quad (2^1 + 2^3 + 2^8 = 266).$

"H" "HTH" "— ... —".

What if we want to see longer sequence? System grows computationally complicated.

Trick: Constructing a martingale. (Shuo-Yen Li, 1980)

(Like a lot of other cases in probability).

"Gambler's team".

Setting. Want to determine $\{X_1\ X_2\ X_3 \dots X_m\}$. $X_i \in \{1, 0\}$.
(H, T)

- We open a casino. we are the dealer.

  Each round we toss a coin.

  Outcome: $Z_1, Z_2, \dots$

Each flip, one more player

- Invite players. $P_0, P_1, P_2, \dots$ to bet on my daily activity

  Player i enters the casino just before round $i+1$.

- Each player enters 1kr of wealth. they only leave if

  They lost everything / seen the sequence $\{X_1, \dots, X_m\}$.

- At first round, P_i bets 1kr on $\{Z_{i+1} = x_1\}$.

$$\begin{cases} 2 \times 1, & Z_{i+1} = x_1 \quad (\tfrac{1}{2}) \\ 0, & Z_{i+1} = 1 - x_1 \quad (\tfrac{1}{2}) \end{cases}$$

- 2nd round. if in the game. bets 2 on $\{Z_{i+2} = x_2\}$.

$$\begin{cases} 2 \times Z, & Z_{i+2} = X_2 \quad (\tfrac{1}{2}) \\ 0, & Z_{i+2} = 1 - X_2 \quad (\tfrac{1}{2}) \end{cases}$$

" Fair game " !

After some rounds, someone gets lucky and sees the pattern

$\longrightarrow$ leaves the game.

After k rounds. Net gain: — Before seeing the sequence.

$$\begin{cases} 2^k - 1 & \text{if still in the game.} \\ -1 & \text{, otherwise} \end{cases}$$

Let G_n^i = Net gain of PL i after round n.

$\delta_n^i = \mathbb{1}_{\text{PL } i \text{ in game after round } n}$.

$$\boxed{G_n^i = (2^{n-i} - 1)\delta_n^i + (1 - \delta_n^i) = 2^{n-i}\delta_n^i - 1}$$

My gain $M_n = \sum_{i=0}^{n-1} G_n^i$. (money collected by me up to time t)

Fact. M is a martingale. $M_0 = 0$

$$\mathbb{E}[M_n \mid M_{n-1}, \ldots, M_0] = M_{n-1}$$

(M_n only depends on the history up to $n-1$,

odds are fair for each player).

"Doob's optional sampling theorem"

$$M_0 = E[M_\tau]. \quad \text{for some stopping time } \tau.$$

for τ s.t. $E[\tau] < \infty$. and

$$E[|M_{n+1} - M_n|] < c. \quad \text{a.s.}$$

Define $\tau := \inf\{ n : Z_{n-m+1} = x_1, \dots, Z_n = x_m \}$

<u>Obs</u>. ① $\tau < mT$. where

$$T := \inf\{ n : Z_{mn+1} \cdots Z_{m(n+1)} = [x_1, \dots, x_n] \}$$

$T \sim geo(\cdot)$

$\Rightarrow E[\tau] < \infty$.

② At any time n. at most m people betting

Total increments $< m 2^m$.

$$0 = M_0 = E[M_\tau] = E\left[\sum_{i=0}^{\tau-1} \left(2^{\tau-i} \delta_\tau^i - 1 \right) \right]$$

$$\Rightarrow \quad E[\tau] = E\left[\sum_{i=0}^{\tau-1} 2^{\tau-i} \delta_\tau^i \right].$$

what we want.

<u>Fact</u>: At time τ, RHS is deterministic!

<u>e.g.</u> "HTHT"

At time τ, how many players are still in the game?

Well at most 4 !

	Want	got	δ
$P_{\tau-1}$.	H	T	0
$P_{\tau-2}$.	HT	HT	1
$P_{\tau-3}$.	HTH.	THT	0
$P_{\tau-4}$.	—	—	1.

$$RHS = 2^{\tau-(\tau-2)} + 2^{\tau-(\tau-4)} = 2^2 + 2^4$$

In the general case, $P\!L_{\tau-i}$ is only in the game if the first i letters are the same as the last i letters.

$$E[\tau] = \sum_{i=0}^{\tau-1} 2^{\tau-i} \delta_\tau^i$$

<u>Check</u>.

"HH". $\quad 2^1 + 2^2 = 6$

"HT" $\quad 2^2 = 4$

"HTHHTHTH". $\quad 2^1 + 2^3 + 2^8 = 2 + 8 + 256$

$= 266.$